Correlated Probabilistic Trajectories for Pedestrian Motion Detection

Frank Perbet   Atsuto Maki   Björn Stenger

Toshiba Research Europe, Cambridge Research Laboratory

http://www.toshiba-europe.com/research/crl/cvg

Abstract

This paper introduces an algorithm for detecting walking motion using point trajectories in video sequences. Given a number of point trajectories, we identify those which are spatio-temporally correlated as arising from feet in walking motion. Unlike existing techniques we do not assume clean point tracks but instead propose "probabilistic trajectories" as new features to classify. These are extracted from directed acyclic graphs whose edges represent temporal point correspondences and are weighted with their matching probability in terms of appearance and location. This representation tolerates the inherent trajectory ambiguity, for example due to occlusions. We then learn the correlation between the movement of two feet using a random forest classifier. The effectiveness of the algorithm is demonstrated in experiments on image sequences captured with a static camera.

1. Introduction

Trajectories of points in image sequences provide a strong visual cue, often allowing the human brain to interpret the scene. Point motion not only gives strong cues about the underlying geometry, but may also be characteristic for an object class. For example, when points close to the joints of a walking person are tracked, the psychological effect of *kinetic depth* allows us to perceive walking motion solely from the 2D point motion pattern. This has first been studied by Johansson using moving light displays (MLDs) [18]. The goal in this paper is to achieve this recognition ability for detecting pedestrian motion from tracked points on a pair of feet whose trajectories are characteristic and spatio-temporally correlated. See Figure 1 for an example of such trajectories.

The biological phenomenon has inspired a lot of work in the area of motion-based recognition, for example human gait analysis [9, 13]. These methods typically require a robust method for feature extraction. Thus, trajectories of interest points are either obtained by using markers to simplify image analysis [8] or are acquired from motion captured data [22] in place of MLDs. There has been relatively little work on recognition purely from low-level features extracted from natural image sequences [24]. Obtaining accurate point tracks is difficult in many cases due to effects such as occlusions, lighting changes and image noise [1].

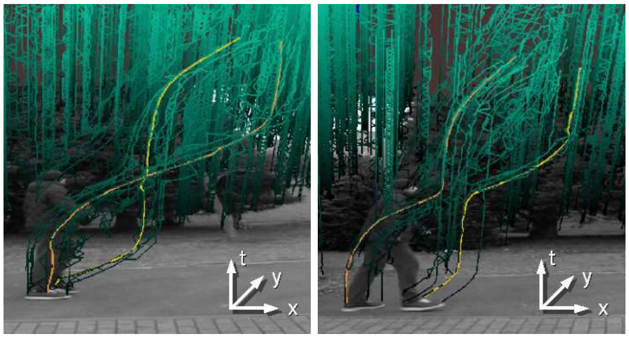


Figure 1. **Space-time volume with point trajectories.** *Brighter green indicates corner positions more recent in time.* **Left:** *Two sample trajectories of corners on the feet are highlighted.* **Right:** *Another case where features are swapped during the short occlusion. Our method is able to correctly classify both cases as walking motion.*

In this application, in order to obtain a discriminative trajectory, a corner should ideally be tracked during a complete walk cycle (about one second). Deterministic point trajectories of typical outdoor scenes are rarely reliable over such a long period of time. This paper does not assume clean point tracks, but retains the concept of temporal connectedness by introducing the notion of *probabilistic trajectories*. These are sampled from a directed acyclic graph whose edges represent temporal point correspondences weighted by their matching probability in terms of appearance and location. Temporal correspondence is thus hypothesized while sacrificing matching accuracy in order to obtain longer and more discriminative trajectories.

The key idea for walk detection from trajectories is to detect *correlated spatio-temporal features*. That is, we detect pedestrian motion by observing the correlation between the motion of two feet of the same person. Recent related

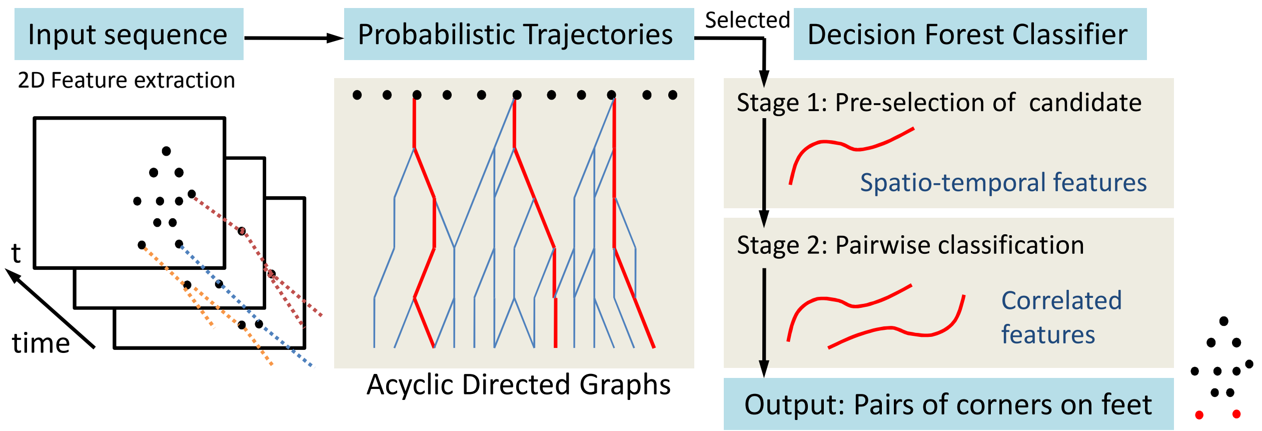

Input sequence
2D Feature extraction
t
time
Probabilistic Trajectories
Acyclic Directed Graphs
Selected
Decision Forest Classifier
Stage 1: Pre-selection of candidate
Spatio-temporal features
Stage 2: Pairwise classification
Correlated features
Output: Pairs of corners on feet

Figure 2. **Schematic of the algorithm.** *Given a video sequence and 2D corners detected in each frame, we first sample probabilistic trajectories of corners in the graph, and then classify the trajectories by a two-stage random decision forest. We design correlated spatio-temporal features for classification.*

work by Brostow and Cipolla includes a method for crowd tracking in which coherent trajectories are found to identify those originating from the same individual [5]. A spatio-temporal volume has been proposed for action recognition by Yilmaz and Shah [27], however it is assumed the object silhouettes to be available. Mikolajczyk and Uemura detect features with different detectors and track them with a KLT tracker [23]. After a motion compensation step actions are recognized using a vocabulary tree. Other related work is that by Laptev and Lindeberg who used space-time interest points[20].

In this paper we choose to use corner features in order to continue detecting points when they are stationary during the walk cycle. Note that physical models of bipedal motion have recently been used in tracking a walking person [7]. We aim to directly learn to detect this type of foot motion in a discriminative manner, while ignoring arm motion which typically exhibits more variation. A further advantage of detecting points on the feet is that their 2D location in the image can directly be mapped to 3D distance given a calibrated camera. This is especially useful in surveillance applications using monocular pedestrian detection [12].

We train a classifier for pedestrian motion of a pair of feet among a number of point trajectories. Intuitively, the motion of a point on a single foot is composed of two periods of dynamic and static phases [2] and motion of points from a pair of feet are alternating in a cyclic manner. In this work we opt for a learning based approach and employ a random forest classifier [4, 14] which has been successfully applied to different classification tasks [6, 21, 25, 26]. It is also well suited to our probabilistic input. That is, we use sampled subgraphs of probabilistic trajectories as input data. We build a two-stage decision forest classifier. The first stage identifies candidate foot trajectories and the second stage associates candidate points as pairs.

The contributions of the paper are thus three-fold: (i) the introduction of *probabilistic trajectories* which temporally associate each point over a sufficiently long time period under both image noise and occlusion, (ii) the pairwise analysis of trajectories for detecting characteristic correlation between the two feet in walking motion, and (iii) the design of efficient features which are computed in the two-staged randomized decision forest classifier. Figure 2 shows a schematic of our algorithm.

The assumptions are that the camera captures dynamics of motion at sufficiently high rate (we use 60 fps) and that people walk with approximately constant speed and direction during a gait cycle. We also assume a stationary camera in this paper, but discuss the extensions for the case of a moving camera.

2. Probabilistic Trajectories

This work uses point trajectories as features for detecting motion of pedestrians. Since repeated detection of the same point over a long time interval is not always possible, we introduce the notion of *probabilistic trajectory*. The basic idea is to hypothesize trajectories by enforcing temporal correspondences between consecutive frames. In practice, corners of two consecutive frames are connected probabilistically using their spatial distance and their appearance. Over T frames, those connections form a graph of possible trajectories. A walk in this graph describes a possible trajectory of a given corner over time, also including many incorrect trajectories. Our assumption is that most of these will still be discriminative, see Figure 1, right. Probabilistic trajectories are generated in three successive steps as described below.

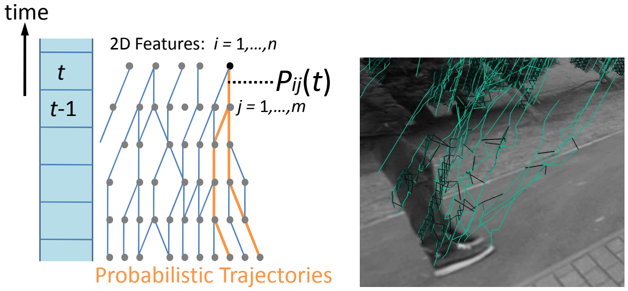


Figure 3. **Probabilistic Trajectories Left:** *A sketch of a graph and selected probabilistic trajectories as our motion descriptor.* **Right:** *An example of partial graph with varying color representing different probabilities, i.e. brighter indicates higher values.*

2.1. Matching Between Two Consecutive Frames

In every frame we extract Harris corners [15] and find potential ancestors for each point among the feature set from the previous frame. Let $p_i(t), i = 1, ..., n$ be the i^{th} corner detected at a 2D location $\mathbf{x}_i(t) \in \mathbf{R}^2$ at time t and let $p_j(t-1), j = 1, ..., m$ be the j^{th} corner found at $\mathbf{x}_j(t-1)$ among m corners which were within a certain range from $\mathbf{x}_i(t)$ in frame $t-1$. We then define the temporal matching score, $P_{ij}(t)$, that $p_i(t)$ matches $p_j(t-1)$ in terms of their appearance similarity S_{ij}, and the spatial distance D_{ij}, by

$$P_{ij}(p_i(t), p_j(t-1)) \propto \exp(-\alpha S_{ij}) \ \exp(-\beta D_{ij}) \ \ , \quad (1)$$

where α and β are positive weighting coefficients.

The appearance similarity S_{ij} is computed from the local image regions around $p_i(t)$ and $p_j(t-1)$, respectively, as the SAD score between them (after subtracting the mean intensity of each image patch) and D_{ij} by their spatial distance $D_{ij} = \|\mathbf{x}_i(t) - \mathbf{x}_j(t-1)\|$.

We represent the existence of a potential match between $p_i(t)$ and $p_j(t-1)$ as a binary value, $E_{ij}(t) \in \{0, 1\}$, based on $P_{ij}(t)$ and define the match as active, $E_{ij}(t) = 1$, with the condition:

$$P_{ij} > \max_j P_{ij} - e \ , \quad (2)$$

where the threshold value e is dynamically adjusted so that the number of pairs is constant (4n). Note that this may result in no active matches for some corners with low values of $\max_j P_{ij}$. We also add temporal matches for the same set of consecutive frames in the forward direction by repeating the process in a reverse manner.

2.2. Acyclic Graph with Matching Probabilities

For each time step t we have determined temporal matches $E_{ij}(t)$ between corners across previous adjacent frames. We retain these for the last T frames (the choice of T will be discussed later). Defining each point $p_i(t)$ as a root node, we generate an acyclic graph, $\mathcal{G}_i(N, E)$, of depth T by tracing active temporal matches along the time axis backward for T frames, see Figure 3. The graph $\mathcal{G}_i(N, E)$ consists of nodes, N, which represent matched corners in the preceding T frames, and edges, E, connecting these nodes. Namely, $E = (E_{ij}(\tau), \tau = t, ..., t-T+1)$. An edge representing an active match, $E_{ij}(t)$, has $P_{ij}(t)$ as its associated weight.

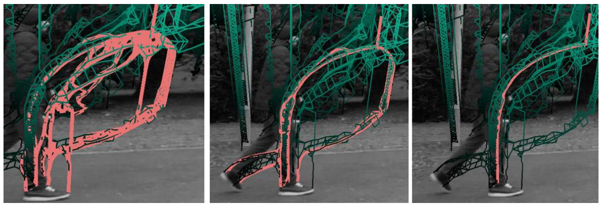


Figure 4. **The influence of the velocity conservation parameter δ. Left:** $\delta = 0.1$ *is small, resulting in flexible but not plausible trajectories.* **Middle:** $\delta = 1$ *is used in this paper.* **Right:** $\delta = 10$ *is high, resulting in a nearly deterministic graph traversal.*

Note that the number of frames, d, for which each corner can be traced back (until encountering an inactive edge or a 'dead end') is available for each node. For example, $d[E_{ij}(t)] = 1$ if $p_j(t-1)$ has no ancestor and $E_{jk}(t-1) = 0$ for all k (where k is an index to features in frame $t-2$). We assign d to each node as its attribute while ideally $d > T$ where at least one path can be found containing nodes from the T previous frames.

2.3. Sampling Probabilistic Trajectories

In each time step the graph is updated and trajectories are sampled from it that are then classified. Intuitively, the sampled trajectories need to be long and physically plausible. Now, we define the *probabilistic trajectories*, $X_i(t) \in \mathbf{R}^{2T}$ of $p_i(t)$, as the paths connecting the root node to different leaf nodes of $\mathcal{G}_i(N, E)$. In practice, a graph traversal of $\mathcal{G}_i$ guided by a probabilistic selection of edges at each node results in plausible trajectories. In particular, we use the sampling probability, $\widehat{P}_{ij}$, in which we also take into consideration the traceable depth d and the velocity conservation factor, V_{ij}:

$$\widehat{P}_{ij}(p_i(t), p_j(t-1)) \propto P_{ij} \exp\left(-\frac{\gamma}{d[E_{ij}]+1}\right) \ \exp(-\delta V_{ij}) \ \ , \quad (3)$$

where γ and δ are positive weighting coefficients, and the last factor

$$V_{ij}(\tau) = \|(\mathbf{x}_h(\tau+1) - \mathbf{x}_i(\tau)) - (\mathbf{x}_i(\tau) - \mathbf{x}_j(\tau-1))\| \quad (4)$$

is valid when $\tau < t$ (so that the coordinate of the previous node in the path, $\mathbf{x}_h(\tau+1)$, is available). We set $\gamma = 10$ and $\delta = 1$ in our experiments. See Fig. 4 for the influence of δ.

3. Classification by Random Decision Forest

Given a corner, $p_i(t)$, and its probabilistic trajectory, $X_i(t)$, our task is now to determine whether or not $X_i(t)$ is the trajectory of a foot during walking motion. In order for a trajectory to contain discriminative features, we consider its length T as roughly covering one walk cycle. As mentioned above, the key idea is to observe point trajectories in pairs. That is, we also consider $p_u(t)(u \neq i)$ that are located in the neighborhood of $p_i(t)$ and examine the spatio-temporal correlation between the probabilistic trajectories, $X_i(t)$ and $X_u(t)$. In order to avoid examining the large number of possible pairs we also use the fact that some trajectories can be rejected immediately as candidates, such as those from stationary background points or those that are too noisy due to incorrect temporal association. Thus, we employ a two-stage classification process:

1. Selection of candidate trajectories.
2. Pairwise classification of pertinent trajectories.

We perform classification using random decision forests in both stages.

3.1. Selection of Candidate Trajectories

The feature design for the first stage is based on the observation that $X_i(t)$ originating from a foot is characterized by dynamic and static phases, being distinguishable from simple trajectories coming from background.

Feature Vectors Let a trajectory, $X_i(t)$, be represented by a vector $X_i(t) = [\mathbf{x}(t), \mathbf{x}(t-1), ..., \mathbf{x}(t-T+1)]^\top$. We first remove its linear component, $\bar{X}_i(t)$, and convert $X_i(t)$ to its canonical form, $\tilde{X}_i(t) = [\tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(t-1), ..., \tilde{\mathbf{x}}(t-T+1)]^\top$. The canonical form, $\tilde{X}_i(t)$, of a trajectory $X_i(t)$, is computed as

$$\tilde{X}_i(t) = X_i(t) - \bar{X}_i(t) \tag{5}$$

where $\bar{X}_i(t) = [\bar{\mathbf{x}}(t), ..., \bar{\mathbf{x}}(t-T+1)]^\top$ and

$$\bar{\mathbf{x}}(\tau) = \frac{1}{T{-}1}[(t{-}\tau)\,\mathbf{x}(t{-}T{+}1) + (\tau{-}t{+}T{-}1)\,\mathbf{x}(t)]. \tag{6}$$

The merit of using the canonical form, $\tilde{X}_i(t)$, is that it represents the motion characteristics independent of its location.

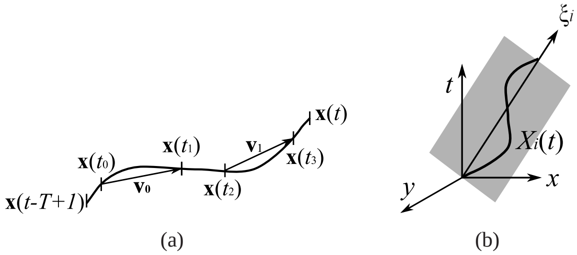
We generate two feature vectors from $\tilde{X}_i(t)$, $\mathbf{v}_0$ and $\mathbf{v}_1$ as the velocity term. By randomly choosing four time instances as cutting points, $t_c(c = 0, ..., 3; t_c < t_{c+1})$, we extract

$$\begin{aligned} \mathbf{v}_0 &= \bar{\mathbf{x}}(t_1) - \bar{\mathbf{x}}(t_0), && (7) \\ \mathbf{v}_1 &= \bar{\mathbf{x}}(t_3) - \bar{\mathbf{x}}(t_2). && (8) \end{aligned}$$

Namely, we sample two random velocities, $\mathbf{v}_0$ and $\mathbf{v}_1$, along a trajectory by choosing two points per velocity, see Figure 5(a). This operation of cutting a trajectory at four points is motivated by the observation that four dynamical models per gait cycle is a reasonable choice in a probabilistic decomposition human gait [3] where coherent motion is used as low-level primitives. We then define our features, f_s and f_d, by the distance and the inner product of scaled versions of the two vectors:

$$\begin{aligned} f_s &= \|a_0\mathbf{v}_0 - a_1\mathbf{v}_1\|, && (9) \\ f_d &= \langle b_0\mathbf{v}_0, b_1\mathbf{v}_1\rangle, && (10) \end{aligned}$$

where a_i and $b_i, i = 0, 1$ are random coefficients in $(0, 1)$. Different features f_s and f_d are generated by sampling values for the coefficients a_i and b_i, as well as the cutting points, $t_c(c = 0, ..., 3)$, of the trajectory.



(a) (b)

Figure 5. **Features from Trajectories.** **(a)** *in order to compute features we sample many pairs of velocity vectors from a trajectory.* **(b)** *The principal direction ξ of the trajectory $X_i(t)$ is used for the directional feature computation.*

Selection of Trajectories We classify candidate trajectories with a random forest [4, 14] which is an ensemble of F decision trees. Each tree examines all input trajectories, $\tilde{X}_i(t)$. Given an input trajectory at the root node, each decision tree recursively branches left or right down to the leaf nodes according to the feature response, f_s and f_d in (10), of a learned function at each non-leaf node. At the leaf nodes, we obtain the class distributions of foot/non-foot. The output from F randomized decision trees is averaged to select candidate trajectories.

Learning Using Random Samples We obtain training data by manually annotating points corresponding to foot regions in a video, and then extracting probabilistic trajectories, $X_i(t)$ of length T, as random subgraphs which stem from the annotated corners. Training is performed separately for each tree using a random subset of the training data.

We recursively split the training data at each node, using the standard method involving information gain [26]. At each leaf node, the class distribution of foot/non-foot is computed from the number of instances that reach the node.

We annotate the ground truth data of feet with a tag of left/right foot so that they can be directly used for training the decision forest in the second stage. It should be noted that those points that can be associated with both feet are also annotated with equal probabilities of being on left/right foot. See Figure 6 for an example of ground truth labels.

3.2. Pairwise Classification of Walking Motion

Given that a corner $p_i(t)$ is selected as a candidate point in the first stage, we pick those $p_u(t)(u \neq i)$ which are located in the neighborhood of $p_i(t)$ and examine how their probabilistic trajectories, $\tilde{X}_i(t)$ and $\tilde{X}_u(t)$, are spatio-temporally correlated.

Features for Directional Correlation Although the trajectory, $X_i(t)$, is three-dimensional, when walking in a straight line, the trajectory lies approximately in a 2D plane. If a set of two candidate trajectories, $X_i(t)$ and $X_u(t)$, arises from walking motion of two feet, the orientations of their 2D planes in 3D space should be close to each other [16]. Based on this observation, we compute the covariance matrices, C_i, of $\tilde{\mathbf{x}}(\tau), \tau = t, ..., t-T+1$, and the eigenvector, $\xi_i \in \mathbf{R}^2$, corresponding to the greatest eigenvalue so that ξ_i represents the principal direction of $\tilde{X}_i(t)$ along its 2D plane, see Figure 5(b). Analogously ξ_u is computed for $\tilde{X}_u(t)$.

We expect ξ_i and ξ_u to be approximately parallel, their directions should be both close to the walking direction. For most of the gait cycle the vector vector connecting the two front points on the trajectory $\mathbf{x}_{iu}(t) = \mathbf{x}_i(t) - \mathbf{x}_u(t)$ can be used as an approximation for this direction. We compute a feature vector containing inner products, $\mathbf{c} \in \mathbf{R}^3$,

$$\mathbf{c} = \begin{bmatrix} \|\langle \xi_i, \xi_u \rangle\| \\ \|\langle \xi_i, \mathbf{x}_{iu}(t) \rangle\| \\ \|\langle \xi_u, \mathbf{x}_{iu}(t) \rangle\| \end{bmatrix} \tag{11}$$

and a random vector $\phi \in \mathbf{R}^3$, $\|\phi\| = 1$, so that

$$f_o \quad = \quad \langle \phi, \mathbf{c} \rangle. \tag{12}$$

Features for Walking Phase Correlation We also design a feature based on the fact that trajectories from a pair of feet are out of phase with each other, alternating in a cyclic manner with dynamic and static phases. This means that one foot is mainly in the dynamic phase while the other is in the static phase. Since one has nearly zero velocity during most of the cycle, we can expect the dot product of their velocity vectors, after proper rectification, to be also close to zero. For this purpose we consider the trajectory, $X_i(t)$, in terms of velocity by generating a vector

$$Y_i(t) = [\mathbf{y}(t), \mathbf{y}(t-1), ..., \mathbf{y}(t-T+2)]^\top \in \mathbf{R}^{2(T-1)} \tag{13}$$

where $\mathbf{y}(\tau) = \mathbf{x}(\tau) - \mathbf{x}(\tau-1), \tau = t, ..., t-T+2$. We convert each $\mathbf{y}(\tau)$ to $\breve{\mathbf{y}}(\tau)$ by projecting it to the axis of ξ_i.

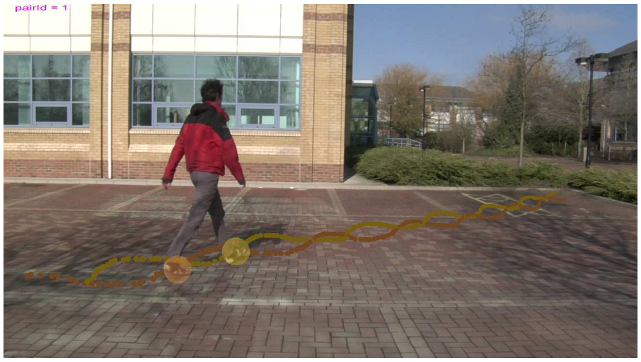


Figure 6. **Ground truth labels:** *Corners detected inside the circles are annotated as being on a foot.*

Thus, the rectified velocity vector is

$$\breve{Y}_i(t) = [\breve{\mathbf{y}}(t), \breve{\mathbf{y}}(t-1), ..., \breve{\mathbf{y}}(t-T+2)]^\top. \tag{14}$$

Rather than simply taking the inner product of the entire $\breve{Y}_i(t)$ and $\breve{Y}_u(t)$, which would result in a scalar, we compute their piecewise dot products. By cutting each of $\breve{Y}_i(t)$ and $\breve{Y}_u(t)$ into l pieces at common fixed cutting points, $t_c(c = 0, ..., l-2; t_c > t_{c+1})$, we acquire a vector

$$\mathbf{q} = [\langle \breve{Y}_i'(t), \breve{Y}_u'(t) \rangle, ..., \langle \breve{Y}_i'(t_{l-2}), \breve{Y}_u'(t_{l-2}) \rangle]^\top \in \mathbf{R}^l, \tag{15}$$

where $\breve{Y}_i'(t_c)$ represents a portion of $\breve{Y}_i(t)$ starting at t_c. We then define a phase feature f_p as the inner product of $\mathbf{q}$ with a random vector, $\psi \in \mathbf{R}^l$ where $\|\psi\| = 1$:

$$f_p \quad = \quad \langle \psi, \mathbf{q} \rangle. \tag{16}$$

We choose to use $l = 5$, again assuming that the four dynamical models per gait are well covered in the trajectories.

Final Detector Output The output from the second decision forest consists of a set of hundreds of feature pairs along with their probabilities of being a pair of feet (see bottom-right in Figure 7 for an example). In order to extract a single pair from this set, we run mean-shift clustering and take the average of the most probable cluster as the final estimate.

4. Experiments

We captured video sequences (resolution 1280×720 pixels at 60 fps) in which a person walks in seven different directions as well as other sequences including different persons walking at different speed.

Figure 7 illustrates the performance of the proposed classification on a sequence of 350 frames. The selected candidates of the first stage and the classified pairs in the second stage are shown in green, in the top and the middle rows, respectively. Although there are some connections between

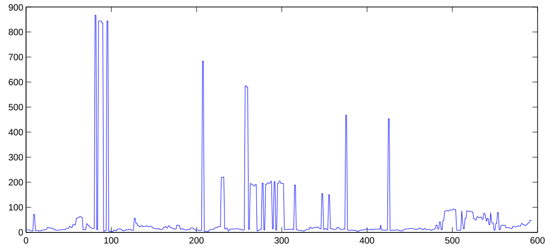


Figure 10. **Estimation error:** *Average distances between detected and annotated pairs plotted for each frame. Peaks in the graph correspond to missed detections of feet.*

corners for example on arms as their motion is similar to feet, the connections between feet are generally dominant, and the final detection results are shown in red in the bottom row. The algorithm currently runs at 2-5 frames per second.

Figure 8 shows an example sequence of 500 frames where outliers become more dominant and the final detection is no longer at the foot location. Although pairs are detected on the feet (left), a number of pairs remain candidates as the result of classification stages. Outliers include pairs of points on the arms as well as pairs between the body and the background.

Figure 9 shows detection in a 400-frame sequence of two pedestrians crossing the scene. During the crossing phase only one pair of feet is detected, but subsequently both are detected again correctly.

For an outdoor sequence with one pedestrian of 595 frames, we compute the error as distance between the detected pairs and the annotated pairs. For this the correspondences between the two pairs is found and the error defined as the average distance between corresponding points, see Figure 10. The average error is less than 30 pixels from the ground truth for 410 frames. Note that the peaks in the error plot correspond to missed detections. For another sequence, the distance was below the same threshold for 310 frames out of 470 input frames. Approximately half of the error cases were due to incorrect detection of arm motion.

5. Discussion

We have introduced a new algorithm for detecting pedestrian motion from point trajectories. In particular, we proposed to use the fact that trajectories from two feet are spatio-temporally correlated. To this end, we have proposed (i) the notion of *probabilistic trajectories* which temporally associate each point over a sufficiently long time period under both image noise and occlusion, (ii) the pairwise analysis of trajectories for detecting correlation between the two feet, and (iii) the design of efficient features for a two-stage random forest classifier. To our knowledge this is the first attempt to recognize walking motion by way of investigating correlation of point trajectories.

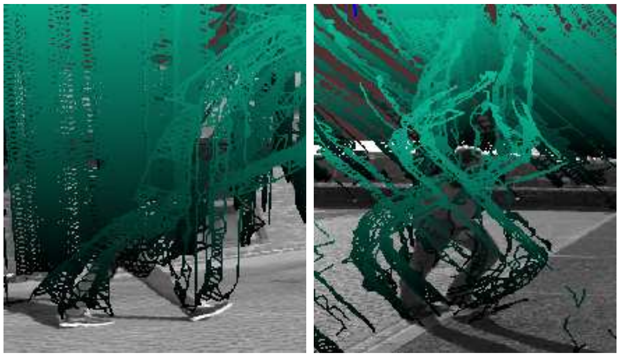


Figure 11. **Static and moving camera.** *Point trajectories in the case of a stationary camera (left) and a moving camera (right).*

The strategy in this paper was to retain the uncertainty in the track associations and let the classifier handle this uncertainty. However, better methods for association could result in less ambiguous trajectories and thus allow the features to be more discriminative. One solution could be to generate shorter but reliable ‘tracklets’ and to link them in an additional step [17].

Future work will also focus on the case of a moving camera. Figure 11 shows trajectories for the cases with both stationary and moving camera. In the case of a stationary camera (left), trajectories connecting background corners are vertically aligned. On the other hand, trajectories viewed by a moving camera (right) exhibit more variation. For sufficiently smooth camera motion it seems that the trajectories of points on the feet are still recognizable. Further options are to employ a global motion compensation step similar to [23] or the *motion features* computed from 3D trajectories introduced in [6].

It will be also useful to model the period of a walk cycle [10, 19]. Currently the method assumes little variation in walking speed.

Finally we point out that the proposed motion-based technique may complement appearance-based methods such as the pedestrian detectors in [11].

Acknowledgments The authors would like to thank Roberto Cipolla for valuable discussions and the reviewers for constructive comments.

References

[1] C. BenAbdelkader, L. S. Davis, and R. Cutler. Motion-based recognition of people in eigengait space. In *FGR*, pages 267–274, 2002.

[2] A. Bissacco. Modeling and learning contact dynamics in human motion. In *CVPR (1)*, pages 421–428, 2005.

[3] C. Bregler. Learning and recognizing human dynamics in video sequences. In *CVPR*, pages 568–575, 1997.

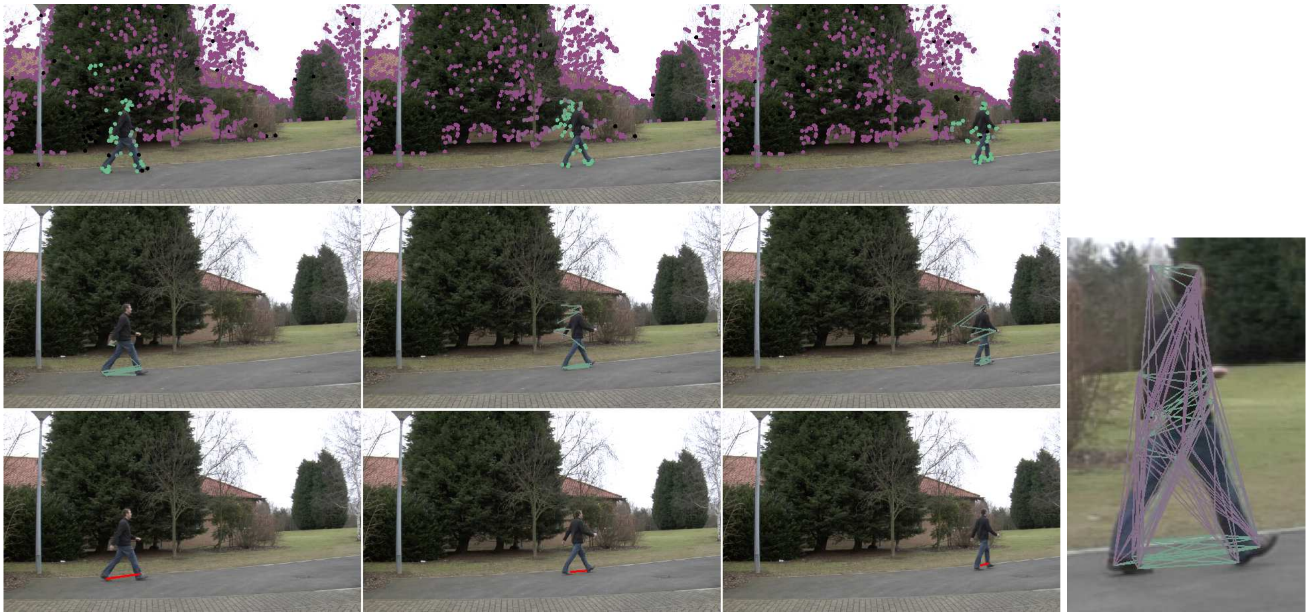


Figure 7. **Results on Sequence I.** *Detection results superimposed on the input frames (from left to right, 150 frames between each view).* **Top:** *Extracted corner points and candidate points after the first stage are shown in green, the rejected points in purple, and points with trajectories of insufficient length in black.* **Middle:** *Pairs of corners extracted as feet in the second stage shown as green line segments.* **Bottom:** *Final detection after running mean-shift.* **Right:** *Example of all possible pairs between pre-selected corners at stage* 1*. Purple line segments are those rejected in stage* 2*.*

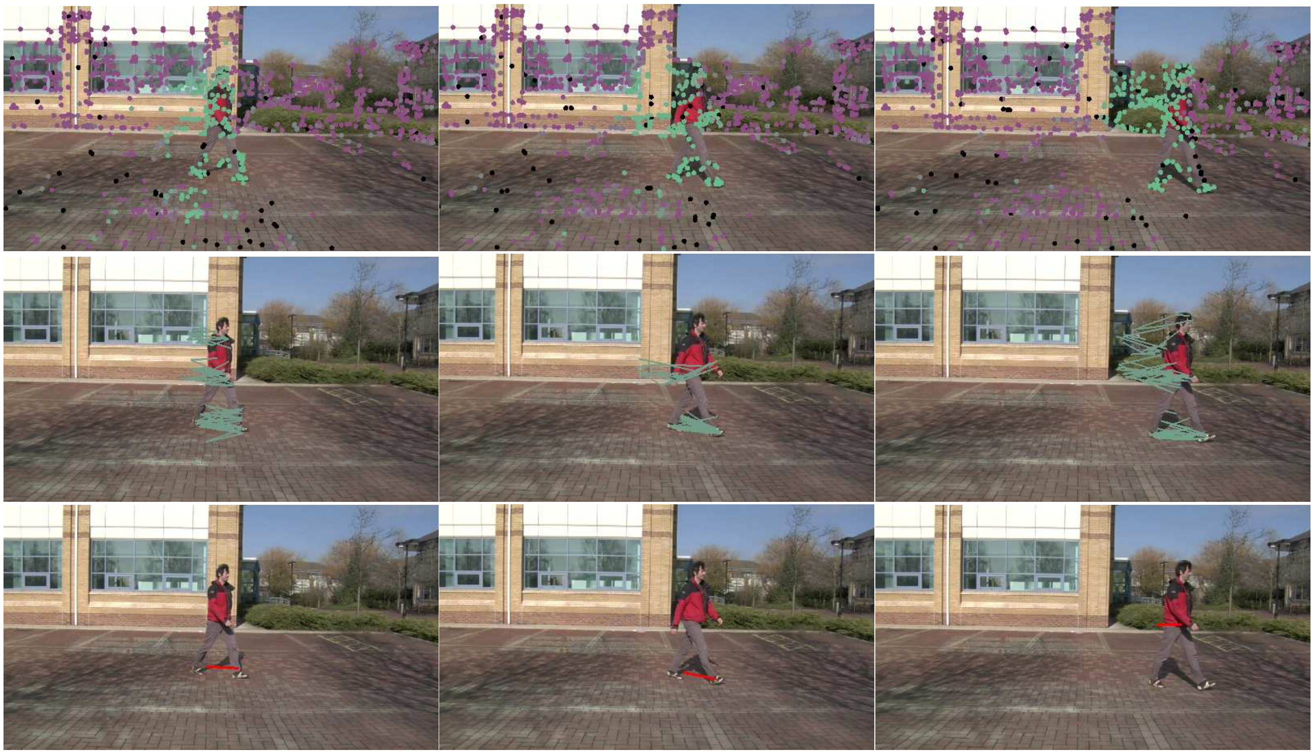


Figure 8. **Results on Sequence II (showing a failure case).** *Feet are correctly detected in the first frames (left, middle), but in the last frame (right) a detection in the arm region occurs due to very similar motion.* **Top:** *Candidate points after the first stage are shown in green, rejected points in purple, points with short trajectories in black.* **Middle:** *Candidate pairs after the second stage are green line segments.* **Bottom:** *Final detection after running mean-shift.*

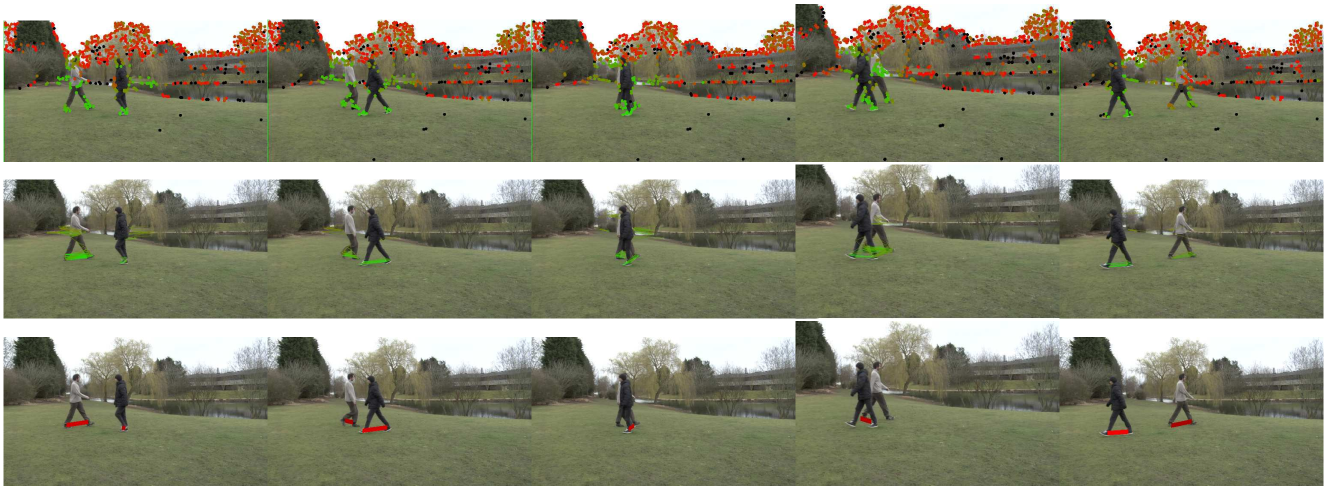


Figure 9. **Results on Sequence III.** *Detection of two pedestrians walking across the scene from opposite directions. During the crossing phase only one pair of feet is detected, but subsequently two pairs are detected again.* **Top:** *Candidate points after the first stage are shown in green, rejected points in red, points with short trajectories in black.* **Middle:** *Candidate pairs after the second stage shown as green line segments.* **Bottom:** *Final detection result shown as red line segments.*

[4] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[5] G. J. Brostow and R. Cipolla. Unsupervised bayesian detection of independent motion in crowds. In *CVPR (1)*, pages 594–601, 2006.

[6] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla. Segmentation and recognition using structure from motion point clouds. In *ECCV (1)*, pages 44–57, 2008.

[7] M. A. Brubaker and D. J. Fleet. The kneed walker for human pose tracking. In *CVPR*, 2008.

[8] L. Campbell and A. Bobick. Recognition of human body motion using phase space constraints. In *ICCV*, pages 624–630, 1995.

[9] C. Cédras and M. A. Shah. Motion based recognition: A survey. *Image and Vision Computing*, 13(2):129–155, 1995.

[10] R. Cutler and L. S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):781–796, 2000.

[11] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR (1)*, pages 886–893, 2005.

[12] M. Enzweiler, P. Kanter, and D. Gavrila. Monocular pedestrian recognition using motion parallax. In *Intelligent Vehicles Symposium*, pages 792–797, 2008.

[13] D. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999.

[14] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 36(1):3–42, 2006.

[15] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference*, pages 147–151, 1988.

[16] D. D. Hoffman and B. E. Flinchbaugh. The interpretation of biological motion. *Biol. Cyb.*, 42:3:195–204, 1982.

[17] C. Huang, B. Wu, and R. Nevatia. Robust object tracking by hierarchical association of detection responses. In *10th ECCV*, volume II, pages 788–801, October 2008.

[18] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2):201–211, 1973.

[19] I. Laptev, S. J. Belongie, P. Pérez, and J. Wills. Periodic motion detection and segmentation via approximate sequence alignment. In *ICCV*, pages 816–823, 2005.

[20] I. Laptev and T. Lindeberg. Space-time interest points. In *ICCV*, pages 432–439, 2003.

[21] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In *CVPR (2)*, pages 775–781, 2005.

[22] Q. Meng, B. Li, and H. Holstein. Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach. *Image Vision Comput.*, 24(8):795–809, 2006.

[23] K. Mikolajczyk and H. Uemura. Action recognition with motion-appearance vocabulary forest. In *CVPR*, pages 1–8, Anchorage, June 2008.

[24] R. Polana and A. Nelson. Low level recognition of human motion (or how to get your man without finding his body parts). In *Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects*, pages 77–82. Press, 1994.

[25] G. Rogez, J. Rihan, S. Ramalingam, C. Orrite, and P. H. S. Torr. Randomized trees for human pose detection. In *CVPR*, 2008.

[26] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *CVPR*, 2008.

[27] A. Yilmaz and M. Shah. A differential geometric approach to representing the human actions. *Computer Vision and Image Understanding*, 109(3):335–351, 2008.